THE IMMERSIVE AUDIO ENVIRONMENT – IMPLEMENTATION, SUBJECTIVE TESTS AND RESULTS

Pratik Shah, Ayman Faza, Raghavendra Nimmala, Steven Grant

Missouri University of Science and Technology
301 W 16th Street, Rolla, MO 65401 USA
{pvsww4, azfdmb, jn8xd, sgrant}@mst.edu

William Chapin

241 Polaris Avenue
Mountain View, California 94043-4514 USA
wchaping@ausim3d.com

ABSTRACT

The effectiveness of virtual environments depends largely on how efficiently they recreate the real world. In the case of auditory virtual environments, the importance of accurate recreation is enhanced since there are no visual cues to assist perception, as in the case of audio-visual virtual environments. In this paper, we present the Immersive Audio Environment (IAE), an easily constructible and portable structure, which is capable of 3-D sound auralization with very high spatial resolution. A novel method for acoustically positioning loudspeakers in space, which is required by the IAE for simulation of sound sources, is presented in this paper. We also present a method to calibrate loudspeakers in an audio system in the case when the loudspeakers are not of the same type. Our contribution is the creation of a system that uses existing and modifications of existing techniques; Vector Based Amplitude Panning (VBAP), in order to recreate an audio battle environment. The IAE recreates the audio effects of battle scenes and can be used as a virtual battle environment for training soldiers. The results of subjective tests show a very low error standard deviation for azimuth and elevation angles and a high correlation between user responses and true angles.

1. INTRODUCTION

The difference between “green” and “battle-hardened” war-fighters is the ability to function effectively in stressful operational environments. This ability comes from well-designed training procedures and experience. While the latter is the best teacher, it is nearly impossible to simulate a battle environment and it is not practical to deploy untrained soldiers in real battle fields for training. Although, live training exercises offer superior realism, they come at a very high cost. As a result, simulation-based training has gained momentum in the last decade. Ideally, audio-visual training environments should be used for training exercises in order to improve realism. The most widely used audio-visual virtual reality environment, CAVE [1], requires multiple projectors and screens; as a result this has a very high setup cost. A derivative of the CAVE, ImmersaDesk [2] uses only a single projector, thus reducing the setup cost. Both, CAVE and ImmersaDesk use 3D visualization glasses, hand tracking and head tracking. This adds to the accessories that the trainees have to put on for training. Such accessories often constrain the movements of the trainees. Unfortunately, for successful implementation of visual environments, tracking is essential. Also, it is difficult to use visual environments to simultaneously train multiple personnel, thereby limiting the types of training exercises that can be performed. One recent virtual reality system, VirKopf [3] uses distributed computing in order to be able to cope with the requirements of audio-video rendering. A more pragmatic option is to use an audio virtual environment to recreate the battle experience, at least in the auditory sense.

A variety of sound reproduction systems have been tried in the past. Stereo [4] and surround-sound [5] present excellent images, but are limited to sounds on the horizontal plane and constrain the listener to be front-facing (theatre-type attitude). Quadraphonics and Ambisonics [6] and also 10.2 audio systems can be very good, but similarly limit the user to theatre-type attitudes. The so-called transaural display [7] delivers convincing aural images from all directions, but the listener has to be in a very tightly constrained space. Tracked-transaural displays free the listener somewhat, but the listener’s nose must still be pointed between the loudspeakers. The ultimate loudspeaker display may be wave field synthesis (WFS) [8]. The position of the virtual sources is constant for the entire listening area and virtual sources can also be simulated *inside* the listening area. Although WFS appears to be a very good sound reproduction technique, it requires a very high spatial sampling rate and thus a large number of loudspeakers, especially for large enclosures. These loudspeakers need to be very accurately positioned. Also, WFS requires very high computational power. Thus, wave field synthesis is cost prohibitive for many more years.

Most of the aforementioned sound reproduction systems are designed in such a way that either there is almost no flexibility for loudspeaker arrangement or they need a very large number of loudspeakers for accurate sound reproduction. Also, the desired effect is best perceived in an area called the “sweet-spot”. The term “sweet-spot” is appropriate because the desired effects are perceived by the listener in a very small area. This limits the ability of the user to perceive the desired effects while moving within the IAE.

Figure 1 shows the Immersive Audio Environment (IAE). The IAE is a 36’ by 18’ by 15’ structure with hardware and software support for 64 loudspeakers including 4 subwoofers. Inspired by Vector-Based Amplitude Panning (VBAP) [9], AuSIM3D’s Vectsonic system is used to drive these 64 semi-arbitrarily positioned loudspeakers. The Vectsonic system enables the extension of the “sweet-spot” to a much larger

Figure 1: The Immersive Audio Environment (IAE). The black floor carpet is the approximate projection of the "sweet-volume".

region with an arbitrary loudspeaker arrangement. This region where the desired effects are perceived as required by the user is termed the "sweet-volume". This allows the trainees to move around freely within the IAE during the training exercises and still accurately perceive the desired audio effects with respect to direction and magnitude. In addition to that, team-based training exercises can be performed due to the extended "sweet-volume".

The rest of the paper is organized as follows: Sections 2 and 3 describe the hardware and software setups of the IAE. The calibration process for the IAE is described in Section 4. The implementation and applications of the IAE are discussed in Section 5. Section 6 concludes the paper with scope for improvements.

2. HARDWARE SETUP

As mentioned before, the IAE is a relatively portable and easy to construct structure. The construction of the support structure of the IAE takes at most 10 hours (a total of 40 staff-hours). The shape of the support structure is seen in Figure 1 and is used mainly to support the loudspeakers and route the signals to the loudspeakers using specialized cables.

Three types of loudspeakers are used in the IAE viz: 44 Myers MM4-XPs, hereafter called the sats, with an operating frequency range of 200 Hz to 18 kHz; twelve Mackie HR824s, hereafter called the mains, with an operating frequency range of 35 Hz to 22 kHz; and four M-Audio SBX10s, hereafter called the subs with a low-frequency operating range of 20 Hz to 200 Hz. Twelve mains are used at approximately ear level since they cover a very large frequency range. Ideally, we would have liked to use all mains instead of the sats. Unfortunately, the mains are considerably heavy and large in size. This makes it difficult to mount them in overhead positions and would place inordinate stress on the support structure. Owing to these considerations, mains are used only at ear level, whereas, at all other locations, sats are used. The advantage of using the sats is that they are light and thus easy to mount. Also, only one cable is needed to supply power and route data to these loudspeakers. The choice of the subs is mainly application specific. The sounds to be generated in the IAE are battle sounds, which have very large low frequency content. Although, such low frequencies are not audible to the human ear, they are felt by the human body in the form of vibrations. Thus they are placed on the floor and form an integral part of any system when simulating battle scenes.

Presently, the IAE uses 60 channels to effectively recreate any audio scene. Although, VBAP supports arbitrarily placed loudspeakers, in the IAE, we have semi-arbitrarily placed the loudspeakers in order to assist and enhance the triangulation process as described in [9]. The four subs are placed close to the base of the rectangle formed by the vertical columns of the IAE. The remaining 56 loudspeakers are arranged in 6 vertical levels. Twelve mains are uniformly distributed around the structure such that the azimuth difference between each of the mains is approximately 30 degrees. This forms level 2. On level 1, twelve sats are placed on the floor, exactly in between the mains. Thus, the floor sats are each 30 degrees apart in azimuth with a 15 degree offset compared to the mains. Another set of 12 sats, with the same azimuth offset as level 1, is connected at an elevation above the mains that is equal to the elevation of the mains above the floor to form level 3. On level 4, ten sats are connected with approximately the same azimuth as the mains with an equivalent increase in elevation. On levels 5 and 6, eight and two sats respectively, are uniformly distributed to cover the remaining area.

3. SOFTWARE SETUP

3.1. Loudspeaker Localization

As stated earlier, VBAP supports arbitrary placement of loudspeakers. However, the location of the loudspeakers in space has to be known in order to create virtual sound sources [9]. Since the IAE may be constructed anywhere, it is not guaranteed that the surface will be level. Thus, manual measurements of heights may lead to errors. The acoustic calculation of loudspeaker positions facilitates experimenting with a variety of loudspeaker placements. The simplest way to find the position of the loudspeakers is using trilateration. Trilateration is the process of locating points in space using the geometry of spheres and triangles. A variety of methods to perform trilateration have been proposed in the literature. The concept of trilateration is mostly used in wireless networks, especially GPS positioning [10] and sensor node localization [11]. These applications generally face the problem of low signal to noise ratio. As a result, the trilateration algorithms have to be quite robust and thus have considerable complexity. For similar reasons, Thomas and Ros [12] have used non-linear methods of trilateration for localizing moving robots.

We present a simple method for locating sound sources in space. Since our environment is acoustically more controlled, we can afford to use simpler methods for trilateration. Our method uses distance equations to acoustically calculate the position of loudspeakers in space. We use 4 microphones located inside the IAE. The transfer function between a loudspeaker and each of the microphones is measured using binary sequences generated using Golay codes [13]. Given the sampling rate and the speed of sound in air, the number of samples to the direct path in the impulse response corresponds to the physical distance of that loudspeaker from the microphone. While calculating the position of the loudspeaker, the speed of sound for current room conditions and the electronic loopback time is also taken into account. It takes about 4 minutes to position all the loudspeakers in the IAE.

Let (x, y, z) be the co-ordinates of the loudspeaker and (x_i, y_i, z_i) be the position of the i^{th} microphone. The distance d_i of the i^{th} microphone from the loudspeaker is given by the distance equation as

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = d_i^2 \quad (1)$$

Since we are using 4 microphones, there will be 4 such equations. To eliminate the square terms, we take the difference of the distance equation for microphone 1 and the distance equations for microphones 2,3 and 4 to form a system of 3 linear equations with 3 unknowns x, y and z. The solution to the system equations is as given by the following:

$\mathbf{s} = \mathbf{X}^{-1}\mathbf{d}$; where $\mathbf{s} = \left[x, y, z\right]^T$,

$\mathbf{X} = \left[\mathbf{p}_{12}^T, \mathbf{p}_{13}^T, \mathbf{p}_{14}^T\right]^{\mathrm{T}}$ and $\mathbf{d} = -0.5 * \left[q_{12}, q_{13}, q_{14}\right]^T$

$\mathbf{p}_{rk} = \left[\left(x_r - x_k\right), \left(y_r - y_k\right), \left(z_r - z_k\right)\right]^T$ is the distance vector for distance betweens microphones r and k

and $q_{rk} = d_r^2 - d_k^2 - \alpha_{rk}$

where $\alpha_{rk} = (x_k^2 - x_r^2) + (y_k^2 - y_r^2) + (z_k^2 - z_r^2)$

3.2. Triangulation

Originally, VBAP uses loudspeaker triplets to create virtual sources anywhere on the plane formed by the three loudspeakers. With 60 loudspeakers, a very large number of loudspeaker triplets can be formed. From all the possible triplets, a small number is selected for simulating virtual sources in the IAE. This process is known as triangulation and is described in [9]. Oftentimes, after triangulation, there are obtuse-angled triangles. These are triangles such that one of the angles is greater than 90 degrees. These triangles are not ideal for VBAP as they result in unequal energy distributions for the three loudspeakers forming the triangle. In such cases, a virtual node is manually added in order to make the triangles more uniform.

Figure 2 shows the advantage of using a virtual node. Before adding the virtual node, triangle 3 on the left has angles significantly larger than 90 degrees. After adding the virtual node, as shown in the figure on the right, all the triangles become relatively more uniform. according to the figure on the left, this virtual node actually lies in triangle 2. If the virtual node is used to simulate any virtual source, the virtual node itself is simulated from the gains obtained from the triangulation on the left. This results in a much more uniform distribution of energies in loudspeakers.

In the current setup of the IAE, triangulation was performed on three sets of loudspeakers. The first set uses four subs and two virtual nodes. The second set includes 12 mains and two virtual nodes. These two sets are used to play low frequency content. The cut-off on the low-pass filter implemented within the subs is set to 50 Hz while the lower cut-off of the band-pass filter implemented within the mains is set to 47 Hz. The third set uses 12 mains and 44 sats for the triangulation process. This set is used to play frequencies above 200 Hz due to the limitation on operating frequency of the sats. The first set consisting of the subs plays frequencies up to 50 Hz. Since the lowest frequency played by the mains is set to 47 Hz, a low-pass filter with the cut-off frequency set to 200 Hz serves the requirement. The third set uses two different types of loudspeakers which are matched by applying a 200 Hz high-pass filter to the virtual channels corresponding to these loudspeakers. To summarize, the first set plays frequencies up to 50 Hz, the second set plays frequencies from 47 Hz to 200 Hz and the last set plays frequencies above 200 Hz. We do not mention the highest frequencies played because they are in the range which the human ear cannot use for localization. The three sets of triangulations result in 4, 12 and 56 virtual channels, in that order. The total number of virtual channels thus becomes 72. These are mixed in order to match the 60 hardware channels using AuSIM3D's Channel Mixer. We believe that this improves the precision with which virtual sources are synthesized.

3.3. Channel Mixing

In the most recent setup, all the mains are used in two sets of triangulations thereby resulting in 2 virtual channels per Mackie and 24 virtual channels corresponding to the mains. These two virtual channels are responsible for different frequency bands. One contains frequencies up to 200 Hz while the other contains frequencies above 200 Hz. A simple addition of these two virtual channels is sufficient as long as the crossover filters are designed with care. AuSIM3D's Channel Mixer is used to perform this patching of virtual channels. As a result of using the Channel Mixer, each virtual channel corresponding to a particular loudspeaker can be patched to its corresponding hardware channel. The mixer also has capabilities to vary the

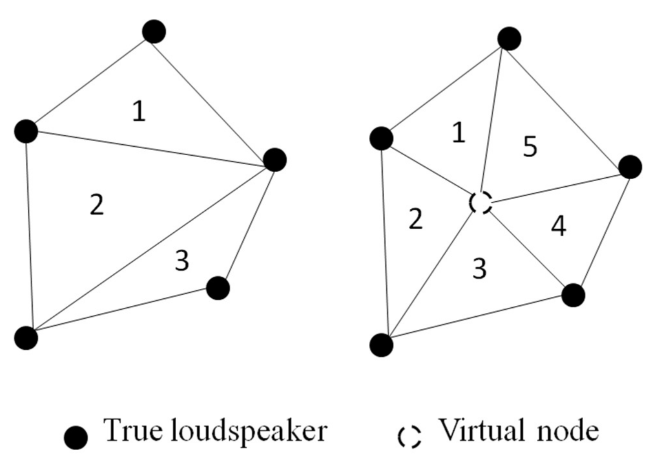

Figure 2: Effect of using virtual node.

gains of virtual channels before mixing. This is useful as two virtual channels have overlapping frequency bands. This facilitates the re-use of loudspeakers for different triangulations thereby enhancing the sound reproductions capabilities of the IAE.

3.4. Creation of AHM file

After successful triangulation, amplitude levels for loudspeaker triplets are calculated for azimuth-elevation pairs with a 3 degree spatial resolution. This choice of the 3 degree resolution is based on the fact that human ear cannot localize within this region. The points for which the levels are not calculated, are interpolated in real-time. The responses from each loudspeaker to each of these positions are also calculated. The responses and the levels are stored in an Acoustic Head Map (AHM) file for rendering with the Vectsonic system. The AHM file also contains the channel patching information from the triangulation process. During the triangulation process, a reference listener position is set. Although, all amplitude levels are calculated for this reference position, the effective listening area is not affected. This information is also contained within the AHM file.

3.5. Sequence Player

The Vectsonic system plays sounds via a sequence player. A sequence is a set of timed events that occur in the audio scene. A sequence can be considered as analogous to a script in a video scene. All timing, position and sound source information is input to the sequence player. The sequence player synthesizes sounds at these defined locations taking into account the doppler and looming effects.

4. CALIBRATION

As described in Sections 2 and 3, three different types of loudspeakers are used in the IAE. To account for the differences in the properties of these loudspeakers, calibration has to be performed. Crossover filters are implemented in order to match the mains with the subs and sats as required by the triangulation process. Also, level calibration is performed in order to match the loudness of the different types of loudspeakers.

4.1. Crossover filter design and implementation

FIR filters are stable and have a linear phase. The use of such FIR filters is ideal for filtering operations in real-time systems with extremely large processing capabilities or ones with very few channels. In our case, a filter is implemented on each channel in real-time. This means 64 filtering operations, one per software channel, have to be performed at a sampling rate of 44.1 kHz. At such high sampling rates, at least 1000 tap FIR filters are required to obtain the required frequency response. The implementation of such long FIR filters requires a lot of processing power and limits the real-time abilities of the system.

The Vectsonic system implements crossover filters as cascaded second order sections of IIR filters. Six cascaded second order sections of IIR filters, designed using Matlab's Filter Design and Analysis Tool, are used to provide the required crossover frequency response. Although the sats have a lower cut-off of 200 Hz, high-pass crossover filters are applied to the sats as well in order to maintain coherency of any sound source synthesized using a combination of the mains and the sats.. For similar reasons, any signal to be played out by the subs is first filtered using a 200 Hz low-pass crossover filter designed for the mains. Figure 3 shows the system level implementation of the crossover filter. The low-pass and high-pass filter for the subs and the sats respectively are implemented on the loudspeakers in hardware.

4.2. Level Calibration

In most systems, the amplitude levels of loudspeakers are calibrated such that the energy received at a certain location, generally the "sweet-spot", from all loudspeakers is the same. This results in a single hot-spot while the other locations have considerably lower amplitude. In our case, since we want to extend the "sweet-spot" to a "sweet-volume" and the loudspeakers are of different types, the calibration process is more complicated.

In order to extend the "sweet-spot" to a "sweet-volume", subsets of loudspeakers are calibrated with spatially distributed microphones, keeping in mind their frequency response. The use of spatially distributed microphones prevents the creation of one bright spot at the center of the IAE and distributes the energies to a larger volume. All loudspeakers present in each sub-section correspond to a subset of loudspeakers. In our case, there are five sub-sections and corresponding to each sub-section is one microphone. All the sats, mains and subs in a particular sub-section are calibrated with the microphone corresponding to that sub-section.

The calibration process starts with calibrating the sats. In order to account for ground and other surface reflections, the transient portion of the received signal is ignored and the energy in the steady state part of the received signal is considered. A white noise sequence generated using Matlab is played through the sats after applying the 200 Hz high-pass crossover filter. Ignoring the first few samples corresponding to the transient portion of the response, the energy in the steady state part of the received signal is adjusted to a constant value, E_i, where i corresponds to the sub-section being calibrated, by varying the gain of the channel. All the sats in a sub-section are calibrated using the same white noise sequence in a similar manner to have the same steady state energy. Once the sats are calibrated, the same sequence is used to calibrate the mains. The sequence is filtered using a 200 Hz high-pass crossover filter and played through the mains. The gains on the mains are adjusted in such a way that the steady state energy for the mains is also equal to E_i. All mains in the sub-section are calibrated using the same procedure. Once the mains are "matched" with the sats, the gains on the mains are not altered. If there is a sub-woofer in the sub-section i, the white noise sequence is low-pass filtered using the 200 Hz low-pass crossover filter and played out through the mains. The steady state energy in this case is denoted as F_i. To calibrate the subs, the low-pass filtered white noise sequence is played through the subs and the gain on the subs is adjusted such that the steady state energy for the subs is equal to F_i. Once

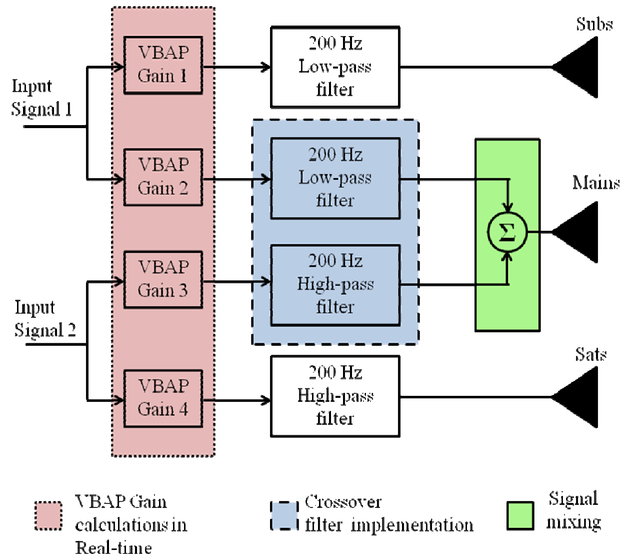

Figure 3: System-level implementation of the crossover filters.

all the gains are set, the system is said to be calibrated and ready for implementation.

5. IMPLEMENTATION AND RESULTS

Presently, only 60 out of the 64 available hardware channels are connected, i.e., only 60 loudspeakers have been used. The triangulation process uses 64 channels. These are "software" channels. The setup is such that more than one "software" channel may map to a single hardware channel, as is the case with the mains. There are 4 mains that are common to both the triangulations. In software, they are treated as distinct channels but when data for these distinct "software" channels has been generated, the data is added together using a mixer and the resultant is 60 hardware output channels.

Audio battle scenes of up to five minutes duration have been created and played out in the IAE. These scenes consist of a spatiotemporal mixture of battle-related sounds such as gun fires, tanks idling, tanks firing, etc. The sounds are synthesized in the IAE using the sequence player. Using the Vectsonic system, the IAE can use up to 64 independent channels to synthesize sounds with a very high spatiotemporal resolution of 3 degrees and a position update frequency of 50 Hz.

A method to evaluate the performance of such systems is presented in [14], where the correlations between the user-observed azimuth and elevation angles with the true azimuth and elevation angles were used as a metric to determine the performance of the system. According to the same study correlations of 0.85, 0.72 and 0.49 were obtained for pairwise amplitude panning, double transaural and ambisonics displays. Based on this study, a system showing a correlation value of 0.85 or higher is considered to be reproducing sounds with excellent accuracy. Unfortunately, correlation is not the best measure of accuracy since an offset in one direction will also result in high values of correlation. To backup a high correlation value, a low value of the difference between the user-observed azimuth and elevation angles and the true azimuth and elevation angles is also required.

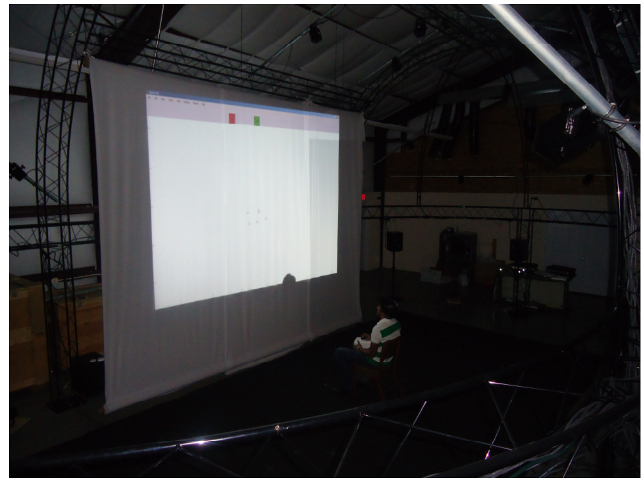

Figure 4: Listener taking the subjective test.

Subjective tests were conducted to test the efficacy of the IAE. Figure 4 shows the experimental setup for conducting the subjective tests. A speaker cloth, which is transparent to sound, is hung in front of the listener and close to the speakers on one side of the IAE. The listener is positioned at the center of the IAE. This is where the amplitude and direction of the synthesized sounds are best perceived. A blank screen is projected onto the speaker cloth. Sounds are generated from behind the speaker cloth spanning the azimuth and elevation covered by the blank screen. The listeners, using a computer mouse, click on the screen in the direction of the perceived sound. The listeners are allowed to change their responses as many times as they wish; only the last response is recorded.

The test for each listener was divided into two parts. The first part was designed to familiarize the subjects with the testing procedure. During this part, sound was played from actual loudspeakers behind the speaker cloth. The sound used for the test was white Gaussian noise sequences generated using Matlab. The listeners clicked on the screen at the point that best described the direction of the perceived sound. As described earlier, the listeners could change their response multiple times. Once the listeners were satisfied with their responses, they confirmed the response and only the final response was recorded. During the second part of the test, the same sound source was synthesized using AuSIM's Vectsonic system. These sounds were synthesized at randomly generated locations within the screen. For each test location, the sound source was played for a duration of four seconds and the listeners were allowed to think for another two seconds before confirming their response. The time constraint was added in order to obtain the worst case response of the listeners. Allowing six seconds for the listener to confirm their responses may not seem like the worst case but this was the minimum time required by the listeners to confirm their responses with some level of confidence. The listeners were also allowed to move their head freely to assist them in the perception of direction. For each subject, a new set of 15

Table 1- Summary of Subjective Test Results

	Test	
	Azimuth	Elevation
Mean	4.7785°	5.2206°
Standard Deviation	3.3606°	4.507°
Correlation	0.9747	0.7818

locations was randomly generated at run-time. As a result, almost all areas of the screen were tested by the time the subjective tests were completed for 12 listeners.

The azimuth and elevation angles for the listeners' responses were recorded. The selection of a reference point for measuring these angles is a difficult task since it may require modeling of the human head. To simplify this task, listeners were asked to position their heads within a small range and the center of this range was considered as the reference point. The range was a volume of a few cubic inches. This does affect the measured angles but this effect is negligible.

Figures 5 and 6 show the best and worst case results of listeners in terms of the correctness of the observed azimuth angle. In each figure, the actual angle (solid line) and the user-observed angle (dashed line) are shown to make the comparison easier. Similarly, Figures 7 and 8 show the corresponding diagrams for the elevation angles.

Table 1 presents a summary of the results obtained from the subjective tests. Specifically, it shows the mean error observed for each angle, and standard deviation, and the correlation between the observed angles and the actual angles. It can be seen that in general, listeners' ability to identify the azimuth angle was generally better than their ability to identify elevation. In addition to that, it can be seen that the correlation between user-observed and actual angles was very strong (0.9747 compared to the results in [14], where any values above 0.85 were considered strong correlation) for the azimuth angles, but was relatively lower (0.7818) for elevation. As for the mean error and standard deviation, it can be seen that in both cases, the results are generally within 6 degrees of error. Given that the listener sits within less than 7 feet from the projection screen, 6 degrees would translate into an error of about 7 inches from the actual location of the sound. Adding to that the fact that the

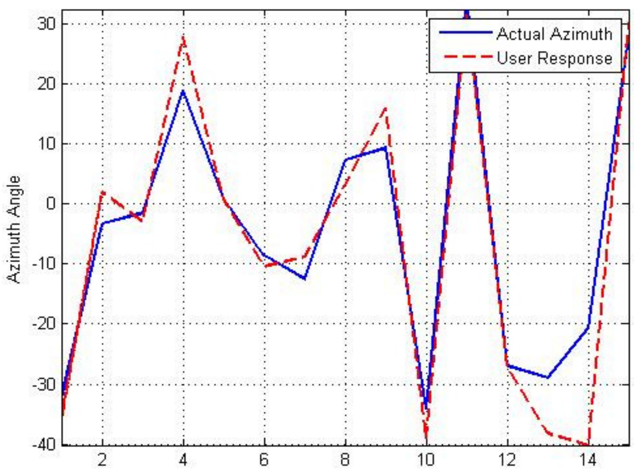

Figure 5: Best Case Azimuth - Observed vs. True Angles

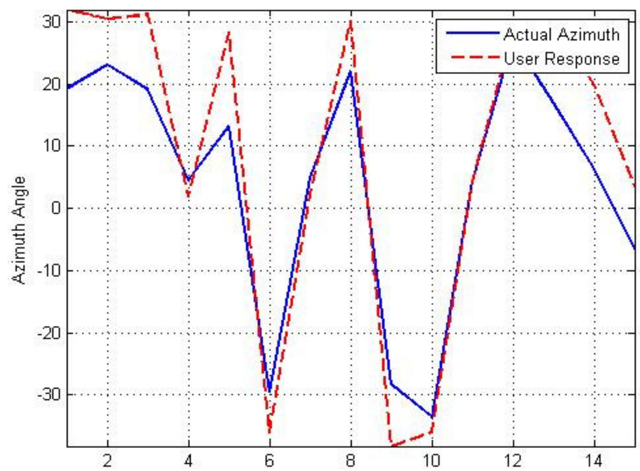

Figure 6: Worst Case Azimuth - Observed vs. True Angles

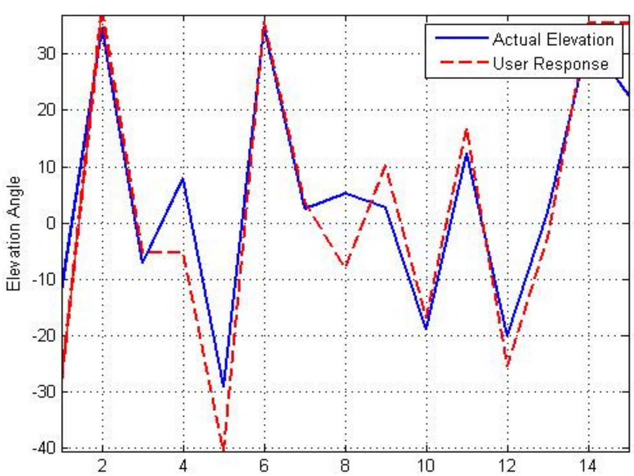

Figure 7: Best Case Elevation - Observed vs. True Angles

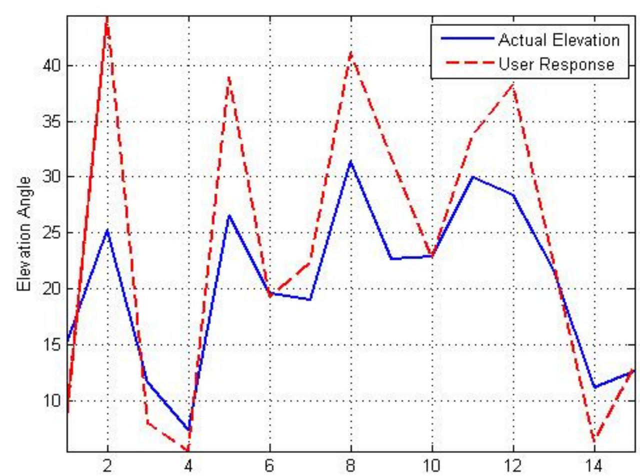

Figure 8: Worst Case Elevation Observed vs. True Angles

human ear can localize sounds up to a 3 degree accuracy in azimuth or elevation, our results show that the system is capable of very accurately generating sounds at specific locations around the listener.

In addition to measuring the spatial accuracy with which the IAE synthesizes sounds, tests were also conducted to verify if the listeners are able to distinguish individual sounds from a mixture. The listeners were immersed in a scripted battle scene. Most of the listeners could identify more than 90% of the sounds played in the battle scene with a good sense of direction. Exposure to longer battle scenes resulted in the listeners experiencing stress due to exposure to battle sounds. This is a very natural phenomena experienced by soldiers in battlefields. Thus the IAE is capable of simuating audio battle scenes with very high spatial accuracy and superior realism.

6. CONCLUSION AND FUTURE WORK

The Immersive Audio Environment (IAE) has been introduced in this paper. We have presented a simple sound source localization method using the IAE. This method uses Golay codes to obtain impulse responses. Golay codes are very effective and efficient for impulse response measurements; however, they have some drawbacks as pointed out in [15]. Thus, sine sweeps may be used for impulse response measurements. System calibration methods have been proposed for systems with different types of loudspeakers. Currently, the IAE synthesizes sound with very high spatiotemporal accuracy. To enhance the accuracy of the system, arrangements with more loudspeakers can be used. Different sets of loudspeakers may be selected to obtain more than two sets of triangulations to obtain a better match for the loudspeakers. We have also performed a number of subjective tests and evaluated the efficacy of the system.

Future work includes validating the system by evaluating the ability of soldiers to correctly locate a target sound in the system. The proposed validation method includes designing experiments in which soldiers are asked to shoot at a specific target sound using a laser gun. Target sounds are synthesized in real time at random locations within the IAE. Each gun used in the experiment will be tracked by infrared cameras, and the tracking software will provide both the location of the gun and its orientation at the time of shooting. This information will in turn be processed to determine whether the gun's position and orientation will in fact point to the correct target location or not. The proposed experiments will be conducted both with battle scenes running in the background, and without the battle scenes, to evaluate the effect of being immersed in a combat like environment, as opposed to performing the same tasks without distractions. Performance results from these experiments would provide us with valuable feedback regarding the effectiveness of our system, and will guide future research and further improvements on this project.

7. ACKNOWLEDGMENT

Research was sponsored by the Leonard Wood Institute in cooperation with the U.S. Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-07-2-0062. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Leonard Wood Institute, the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

8. REFERENCES

[1] C. C. Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart, "The cave: audio visual experience automatic virtual environment," *Communications of the ACM*, vol. 35, no. 6, pp. 64-72, February 1992.

[2] M. Czernuszenko, D. Pape, D. Sandin, T. DeFanti, G. L. Dawe, and M. D. Brown, "The ImmersaDesk and Infinity Wall projection-based virtual reality displays," *SIGGRAPH Computer Graphics,* vol. 31, no. 2, pp. 46-49, May. 1997.

[3] T. Lentz, D. Schröder, M. Vorländer, and I. Assenmacher, "Virtual reality system with integrated sound field simulation and reproduction," *EURASIP Journal of Applied Signal Processing*, vol. 2007, no. 1, pp. 187, 2007.

[4] B. B. Bauer, "Some techniques toward better stereophonic perspective," *IEEE Transactions on Audio*, May-June, 1963, p. 89

[5] S. Quackenbush and J. Herre. "MPEG Surround," *IEEE Multimedia, Vol. 12, Issue 4,* pp 18-23, October-December 2005.

[6] M. A. Gerzon and G. J. Barton, "Ambisonic surround-sound mixing for multitrack studios," in *Proc. of the AES 2nd Int. Conf.: The Art and Technology of Recording*, May 1984.

[7] M. R. Schroeder and B. S. Atal. "Computer simulation of sound transmission in rooms". *IEEE Convention on Record,* vol. 7, pp. 150-155, 1963.

[8] A. J. Berkhout, D. De Vries, and P. Vogel. "Acoustic control by wave field synthesis," *Journal of Acoustic Society of America,* vol. 93, pp. 2764–2778, May 1993.

[9] V. Pulkki. "Spatial sound generation and perception by amplitude panning techniques," PhD thesis, Helsinki University of Technology, Espoo, Finland, 2001.

[10] S. Zirari, P. Canalda, and F. Spies, "WiFi GPS based combined positioning algorithm," *IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS)*, pp.684-688, 25-27 June 2010.

[11] C. Xiuzhen, S. Haining, L. Qilian, and D. Hung-Chang Du, "Silent Positioning in Underwater Acoustic Sensor

Networks," *IEEE Transactions on Vehicular Technology,* vol.57, no.3, pp. 1756-1766, May 2008.

[12] F. Thomas and L. Ros, “Revisiting trilateration for robot localization,” *IEEE Transactions on Robotics,* vol. 21, no. 1, pp. 92-101, February 2005.

[13] S. Foster, “Impulse Response Measurement Using Golay Codes,” Trans. ICASSP ’86, pp. 929-932, 1986.

[14] C. Gustavino, V. Larcher, G. Catusseau, and P. Boussard, Spatial Audio Quality Evaluation: Comparing Transaural, Ambisonics and Stereo, in *Proc. of the 13th Int. Conf. on Auditory Display*, Montreal, Canada, 2007.

[15] S. Müller and P. Massarani, “Transfer Function Measurement with Sweeps,” *Jouranl of Audio Engineering Society,* vol. 49, pp. 443-471, 2001.